The stigma of mental illness remains a serious social problem and critical impediment to treatment seeking among diagnosed individuals. Study 1 evaluated explicit attitudes and stereotypes about persons with mental illness relative to persons with physical illness, and also implicit attitudes that lie outside conscious control (using the Implicit Association Test) in a college sample (N = 119). Study 2 extended the evaluation of explicit and implicit biases to a sample diagnosed with mental illness (N = 35) and a healthy control sample from the general population (N = 36). Results demonstrated implicit negative attitudes and beliefs about the helplessness and blameworthiness of mentally ill persons. Interestingly, relatively negative explicit attitudes and biases about the helplessness (though not blameworthiness) of mentally ill persons were also evident. In addition, being a member of the stigmatized group did not result in lower implicit or explicit biases, suggesting that no protective in–group bias exists.

I wanted to tell her that if only something were wrong with my body it would be fine, I would rather have anything wrong with my body than something wrong with my head.

—Sylvia Plath, 1966, p. 193

The authors are thankful for the research assistance provided by Betsy Hernandez, Zachary Patberg, Elaine Neam, and other members of the Cognition and Psychopathology lab at the University of Virginia, and for the feedback provided by Brian Nosek and Stacey Sinclair, as well as comments from anonymous reviewers on an earlier version of this manuscript. Portions of Study 1 were written up for undergraduate theses by Joel Wilson and Elaine Neam at the University of Virginia, and portions of Study 2 were presented in a poster at the 2004 annual meeting of the Society for Personality and Social Psychology. Finally, the authors wish to thank the Charlottesville Albemarle Community Foundation mental health fund for helping to support this research.

Address correspondence to Bethany A. Teachman, Department of Psychology, University of Virginia, P.O. Box 400400, Charlottesville, VA 22904–4400; E-mail: bteachman@virginia.edu.

© 2006 The Guilford Press
The stigma associated with mental illness presents a serious obstacle both for individuals looking for mental health treatment and for diagnosed persons in the community (Komiya, Good, & Sherod, 2000; Link, Struening, Neese–Todd, Asmussen, & Phelan, 2001; Sirey, Bruce, Alexopoulos, Perlick, Freidman, & Meyers, 2001). To better understand the negative evaluations that contribute to this stigma, the present study examines attitudes and stereotypes about persons with mental illness, both that occur within and outside of conscious control, and evaluates whether being a member of the stigmatized group (i.e., having a mental illness diagnosis) is associated with relatively less bias against other persons with mental illness.

Early investigations of stigma followed from Goffman’s (1963) seminal work defining stigma as any aspect of an individual that is deeply discrediting and thereby allows others to discount that individual as “tainted” (p. 3). More recent conceptions of stigma consider four social–cognitive processes that lead to the discrediting of a person with mental illness (Corrigan, 2004). These include (1) cues, such as mental illness labels, symptoms, and social skills deficits, (2) stereotypes, which are knowledge structures or beliefs learned about a social group, (3) prejudiced attitudes, which reflect an evaluative or affective component, and (4) discriminatory behaviors, which are negative actions against the group. The current investigation focuses on evidence for implicit and explicit stereotypes and prejudice in response to the cue of mental illness labels, though we consider all four components to be important interacting contributors to the stigma process. In addition, both “public stigma,” that is the bias from the general public toward persons with mental illness (Study 1), and “self–stigma,” the bias from persons with mental illness toward their own group that follows from internalizing public stigma (Study 2; see Corrigan, 2004) will be investigated.

VIEWS ON STIGMA OF MENTAL ILLNESS

There has been considerable debate about the extent of the problem posed by stigma of mental illness. On one side of the debate, predominantly espoused in the 1970s and 1980s, theorists have argued against the claim that stigma toward persons with mental illness is widespread (e.g., Rabkin, 1984). This argument followed in part from numerous vignette studies that found little influence of mental illness labels on evaluations of diagnosed persons (see review in Link & Phelan, 1999). In fact, a 1980 National Institute of Mental Health workshop chose not to use the term “stigma” in its title, suggesting that it was not justified if “one is referring to negative attitudes induced by manifestations of psychiatric ill-
ness” (Rabkin, 1984, p. 327). In addition, the 1996 General Social Survey (Pescosolido, Monahan, Link, Stueve, & Kikuzawa, 1999) found that while there is still a general negative attitude associated with mental illness, the majority of respondents reported that they believed persons with major depression were competent to manage autonomous treatment decisions. Further, mental illness advocacy groups have made important strides in recent years (see Corrigan & Penn, 1999) as evidenced by the first–ever White House conference on mental health in 1999.

Yet, the assertion that stigma is either not present (Crocetti, Spiro, & Siassi, 1974), or is diminishing over time (Segal, 1978), has also been fundamentally challenged. First, these claims stand in stark contrast to patients’ own recent reports about the experience of being labeled mentally ill (see Wahl, 1999). Fully 78% of respondents in a survey of mental health consumers reported overhearing hurtful or offensive comments about mental illness, and more than a third reported being treated as less competent by others. In addition, even relatively recently, much of the evidence still indicates that the label “mentally ill” makes it more difficult to obtain work and housing, and to gain acceptance from peers and co–workers, regardless of the individual’s behavior (see Farina, 1998). Further, some research suggests that educational programs to reduce stigma have only minimally affected public knowledge, and that diagnosed persons are still responded to with aversion (Socall & Holtgraves, 1992). Many questions remain about how stigma of mental illness has changed over the last 30 years, and how the methodologies used to measure stigma may lead to divergent results.

COMPARING IMPLICIT AND EXPLICIT BIAS

Although stigma of mental illness has been measured using a variety of self–report and behavioral indicators, there has been little investigation into the role that implicit, less strategic or unintentional, processes play in the expression of bias. This is surprising because across a broad range of stigmatized groups, it is now clear that implicit biases may reveal negative attitudes and beliefs toward marginalized groups, sometimes even when these are not explicitly endorsed (see reviews in Greenwald, Banaji, Rudman, Farnham, Nosek, & Mellot, 2002; Nosek, Banaji, & Greenwald, 2002). Interestingly, evidence suggests that at times implicit and explicit biases will be unrelated, and at other times, these biases may be strongly correlated (Nosek et al., 2002). Importantly, these biases may differentially predict discriminatory behaviors. For instance, Bessenoff and Sherman (2000) used a lexical decision task to demonstrate that implicit anti–fat evaluations predicted how far participants chose to sit from an overweight woman, whereas explicit attitudes did not.
Implicit attitudes capture evaluations that lie outside conscious control (and at times outside conscious awareness; Greenwald & Banaji, 1995). There is now abundant evidence that implicit biases against numerous marginalized groups exist even when participants do not consciously wish to endorse negative attitudes (e.g., Nosek et al., 2002; Teachman, Gapinski, Brownell, Rawlins, & Jeyaram, 2003), pointing to their involuntary nature. This discrepancy can occur when individuals do not want to report negative evaluations because of social desirability concerns, or because the biases reside outside conscious awareness, or because automatic biases can influence responding even though a person wishes to be egalitarian. Alternatively, implicit and explicit attitudes may both show bias in the same direction (e.g., both reflect negative attitudes), but still be distinct in that they reflect unique components of the attitude construct. For instance, in the field of anxiety disorders, implicit and explicit evaluations of spiders were unique predictors of avoidance behaviors among persons with spider phobia (Teachman & Woody, 2003).

How might implicit and explicit biases compare and contrast to enhance our understanding of the stigma of mental illness? It is possible that efforts by mental illness advocacy groups have made it less socially acceptable to explicitly endorse stigmatizing attitudes, but as noted above, also evidence arguing against such changes (see Phelan, Link, Stueve & Pescosolido, 2000). Irrespective of the level of explicit bias, popular stereotypes about mental illness may still be common on an implicit level. We are unaware of any published work evaluating implicit biases against mental illness, or comparing implicit and explicit evaluations of persons with mental illness. Given recent efforts to raise public awareness about the stigma, it is expected that people will self-report only minimal negative attitudes due to concerns about social desirability and recognition of the importance of increased tolerance (though as mentioned earlier, it is clear that evidence for self-reported biases varies greatly across studies making self-reported biases difficult to predict). However, the ongoing evidence of discrimination and pervasive negative messages in the media and among the general public suggest we will find robust evidence for biased implicit associations. Further, minimal relationship between the implicit and explicit measures is predicted, given the likely influence of strong social desirability concerns and suggestive findings that these concerns may moderate the association between implicit and explicit indicators (Nosek, 2004).

MENTAL ILLNESS AND IN–GROUP BIAS

We believe the evidence is overwhelming that stigma against mental illness remains a serious problem, and further, that individuals who are a
member of this marginalized group are aware of that stigma (as evident by the large proportion of mental health consumers reporting experiences of bias; e.g., Wahl, 1999). What are the consequences of perceiving this bias? Will individuals diagnosed with mental illness share the negative attitudes and stereotypes present in the general population, even though they are recipients of these biased views? Alternatively, will their membership in the group be associated with lower bias, perhaps because of greater knowledge about and contact with mental illness and other consumers, or because of the clear incentives to judging one’s in–group positively (Brewer, 1979)?

Evidence from other marginalized groups provides mixed results, leading Crocker (1999) to suggest that the consequences of stigma on self-esteem and internalization of the devaluing experienced by others varies greatly as a function of the social context. For instance, African Americans often evaluate African Americans as a group more positively than Caucasians do (e.g., Nosek et al., 2002), but obese persons are frequently found to hold anti–fat views that are equally as negative as persons who are not overweight (Crandall, 1994). Corrigan and Watson (2002) reviewed the relevant literature on stigma of mental illness, and suggested that the biases “may result in significant loss in self-esteem for some, while others are energized by prejudice and express righteous anger. Added to this complexity is a third group: persons who neither lose self-esteem nor become righteously angry at stigma, instead seemingly ignoring the effects of public prejudice altogether” (p. 35).

Thus, it is difficult to predict whether persons with mental illness will have internalized the negative biases about their group, or will instead display a protective in–group bias, showing reduced negativity toward persons with mental illness. Given the widespread discrimination and stereotypes they have been exposed to, as well as the objective negative outcomes that follow a mental illness diagnosis, we expect that some internalization of stigmatizing views will be evident among persons with mental illness, particularly implicitly, since these bias measures are so difficult to strategically control. However, we hypothesize the biases will be slightly lower than those observed in the general population given that both education about mental illness and direct interaction with persons with mental illness have been associated with positive changes in attributions about psychiatric disabilities (Corrigan et al., 2001). In particular, it is expected that diagnosed persons who are in treatment settings with others, will have both greater contact with, and knowledge about, mental illness than the average person. These hypotheses were explored in Study 2.
STUDY 1

Study 1 evaluated the presence of implicit and explicit attitudes and stereotypes toward persons with mental illness among a student sample as an initial attempt to establish measures of automatic associations that would reflect the stigma against mental illness.

METHOD

Participants
One hundred nineteen undergraduate students at the University of Virginia participated in the experiment in partial completion of requirements for introductory psychology courses. The sample was 57% female, mean age was 19 (SD = 1.21, range = 18–27), and 65% were Caucasian (12% African American, 12% Asian, 4% Hispanic, and 6% indicated “other” for their ethnicity).

Materials

Implicit Bias Measures. A computerized version of the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) was used to assess automatic associations to mental illness relative to physical illness. The associations are automatic in that they reflect attitudes that occur outside conscious control, though not always outside awareness. This method has been widely used to assess implicit attitudes and stereotypes and has adequate psychometric properties (Greenwald & Nosek, 2001). The IAT is a response time task that requires participants to classify word stimuli into superordinate categories to reflect the ease with which participants can associate concepts in memory. The task requires items to be classified while two category labels are paired on either side of the screen. Specifically, the task involves comparing the time taken to classify stimuli when paired categories match a person’s automatic associations versus the time taken when paired categories contradict automatic associations. Thus, it is fundamentally a reaction time task that evaluates differences in time required to classify items into more general level categories. (See http://implicit.harvard.edu/implicit/ for more information and a sample test.)

In the current study, we were interested in implicit associations about mental illness. Because the IAT is a relative task, an equivalent comparison category was required. Mental illness is a generally negative concept given that it reflects illness, so attitudes toward physical illness were used for comparison. These categories were compared while being paired with opposing descriptor categories to reflect common evaluations of mentally ill persons. For example, the categories “mental illness” and “bad” were paired at the top left of the computer screen while
“physical illness” and “good” were simultaneously paired at the top right. Participants were told to classify any stimuli that belonged to either the mental illness or bad categories on the left, and any stimuli that belonged to either the physical illness or good categories on the right. The dependent variable was the speed of classification across a series of trials. Following this category pairing condition, the labels were switched and the same categorization task was completed while pairing “mental illness” with “good” (and “physical illness” with “bad”). Thus, for each IAT task, two sets of category pairs were presented simultaneously.

It is expected that when categories are paired to match a person’s automatic associations, they will be able to classify the stimuli more quickly. Thus, IAT effects are determined by contrasting average response time in one category pairing with average response time in the other. Given the pervasiveness of negative attitudes toward mental illness in Western culture, response times were expected to be faster when the “mental illness” category was paired with negative descriptors (such as “bad”), than when “physical illness” was paired with the same negative descriptors. Participants completed three different IAT tasks (one attitude and two stereotype measures). Each contrasted “physical illness” and “mental illness,” with either: (1) “bad” versus “good” (referred to as “IAT mental illness + bad” because positive values reflect faster responding when mental illness was associated with bad), (2) “blameworthy” versus “innocent” (referred to as “IAT mental illness + blameworthy”), and (3) “helpless” versus “competent” (referred to as “IAT mental illness + helpless”). In each case, it was hypothesized that mental illness would be implicitly evaluated more negatively than physical illness.

Four items were selected for each category. To select stimuli, a list of ten common mental and ten common physical illnesses were pre–tested. An independent sample (N = 20) rated each illness for ease to classify the word into the category and the perceived degree of interference in functioning and level of distress the illness causes. Only illnesses rated 5 or higher for ease to classify (on a 7–point scale) were included. Mental illnesses were then matched to physical illnesses with comparable means for degree of distress and interference. A paired–samples t–test established that the average negativity of the physical and mental illness stimuli did not differ (t_{19} = 1.01, p > .10, d = .46). See Appendix for IAT category labels and stimuli.

In each IAT task, there were two critical trial blocks: one block of trials where the target and descriptor categories reflected negative mental illness associations and one block in which the categories reflected negative physical illness associations. Each critical block consisted of 56 clas-
sification trials; the first 20 were practice (and were excluded from analyses), and the remaining 36 constituted the experimental data.

Explicit Bias Measures. Participants rated their attitudes toward “persons with mental illness” and “persons with physical illness” on 7-point semantic differential scales (1 = bad to 7 = good, referred to as “Explicit mental illness + bad”). Analogous scales were completed for the blameworthy/innocent and helpless/competent stereotypes. These self-report items were designed to parallel the relative nature of the IAT tasks to permit implicit/explicit comparisons. A difference score was calculated for each target attitude, with a positive score indicating a negative evaluation of persons with mental illness relative to physical illness.

In addition, participants also completed the 8-item Perceived Dangerousness Scale (PDS) that includes items related to the expected threat from persons with mental illness. Inclusion of this scale followed findings from Link, Cullen, Frank, and Wozniak (1987) that beliefs about the perceived dangerousness of mentally ill persons moderated evaluations of job applicants labeled as mentally ill. It was hypothesized that this common stereotype would be positively related to both the implicit and explicit bias measures in the current study.

Procedure
Following informed consent, participants completed the three IAT tasks, plus an initial practice task, and the explicit bias measures. Order of the IATs versus questionnaires was counterbalanced, order of the questionnaires within each packet was randomized, and order of the IAT blocks (i.e., bad/good, helpless/competent, blameworthy/inno-

---

1. Other explicit measures were included in the study for exploratory purposes, and are not reported here because they were not part of the study’s central hypotheses. These included various demographic variables, a measure of Devaluation/Discrimination (Link, Cullen, Streuning, Shrout, & Dohrenwend, 1989), the Community Attitudes toward Mentally Ill—en=olence subscale (Taylor & Dear, 1981), the Secrecy subscale from the coping orientations described in Link et al. (1989), and the first author (BT) devised a measure to assess beliefs about the Causes of Mental Illness.

In addition, participants read a vignette about a job applicant that manipulated information about the applicant’s illness (as either physical or mental) and the applicant’s behavior (regarding number of reported negative behaviors), and then evaluated the likelihood of hiring the applicant and desired social distance from the applicant. This part of the protocol was part of a separate study and is not reported here since it was not expected to impact upon the bias measures given that the job applicant manipulation was about a particular person, whereas the implicit and explicit bias measures asked about global evaluations of mentally ill persons. As predicted, the vignette manipulation did not significantly affect the multivariate tests for the implicit (IAT tasks) and explicit (semantic differentials) bias measures (p > .10).
RESULTS

Data Reduction. The IAT data were scored according to the new scoring algorithm developed by Greenwald, Nosek, and Banaji (2003), because this approach maximizes the relationship between implicit and explicit measures, and improves the psychometric properties of the tool (by taking into account each respondent’s latency variability, and including a latency penalty for errors). This resulted in deletion of two participants’ IAT data according to Greenwald et al.’s revised criteria (based on either high error rates above 40% on a critical IAT block and/or unusually fast or slow response times, such as more than 10% trials with latency less than 300 ms; Nosek, 2004, personal communication). The overall IAT error rate was 7% for the final sample, and reliability of the IATs was adequate for response latency data: split–half reliability for average of the three IATs = .58 (individual IATs ranged from $r = .32$ to $.56$).

Evidence of Implicit and Explicit Bias. Positive IAT effects reflect faster response times when mental illness was paired with a negative descriptor and physical illness with a positive descriptor; hence, the effects can be interpreted as both bias against mental illness and/or bias in favor of physical illness. As expected, there was strong evidence of an implicit bias against mental illness relative to physical illness. T-tests indicated that the implicit measures were each significantly different from zero: IAT mental illness + bad ($t_{116} = 4.60, p < .0001, d = .85$); IAT mental illness + helpless ($t_{116} = 5.47, p < .0001, d = 1.02$); and IAT mental illness + blameworthy ($t_{116} = 2.50, p = .01, d = .46$).

Although minimal explicit bias was anticipated on the semantic differential scales, there was evidence of relatively negative Explicit mental illness + bad ($t_{118} = 3.36, p = .001, d = .62$) and Explicit mental illness + helpless ($t_{118} = 5.09, p < .0001, d = .94$). There was no significant effect for Explicit mental illness + blameworthy ($t_{118} = 1.31, p = .19, d = .24$; in fact, the mean was in the opposite direction). For easier visual inspection, Figure 1 indicates the z–scores (with a meaningful zero point) and standard error bars for the implicit and explicit bias measures.

It should be emphasized that the biases were relative (i.e., evaluation of persons with mental compared to physical illnesses). On the explicit measures, where evaluations of the different groups can be teased apart (which is not possible on the implicit tasks), the ratings for each group separately did not indicate negativity for the majority of items (only the rating of mentally ill people as helpless versus competent had a mean below the neutral point on the 7–point scale). Thus, the groups were not...
explicitly evaluated negatively relative to positively, but persons with mental illness were evaluated more negatively than persons with physical illnesses on certain explicit indicators.

Relationships Among Bias Measures and Individual Differences. As expected, the bias measures were not significantly related to one another, suggesting the independence of the constructs (though the relatively low reliability of reaction time measures may also partly account for the low correlations). The implicit tasks were not related to one another ($r$ ranged from $-.03$ to $.17$, all $p > .05$), nor were the explicit tasks interrelated ($r$ ranged from $-.06$ to $.17$, all $p > .05$). In addition, there were no significant gender differences on the bias measures (all $p > .10$). As expected, the Perceived Dangerousness Scale was positively related to Explicit mental illness + helpless ($r = .19$, $p = .04$), but not to the other bias measures ($r$ ranged from $-.04$ to $.13$, all $p > .10$), providing only partial support for the expected positive associations.

STUDY 2

Study 1 established strong evidence for implicit biases against persons with mental illness as bad, helpless, and blameworthy relative to persons with physical illness. Interestingly, there was also evidence of relatively negative explicit evaluations of mentally ill persons as bad and helpless (though not evidence for a negative evaluation of the group when rated separately). As expected, the implicit and explicit bias measures were not strongly related to one another, suggesting they may reflect different components of the stigma of mental illness. Study 2 was designed to replicate the findings of implicit and explicit stigma, and extend the investigation to a sample diagnosed with mental illness to evaluate whether persons who are members of the stigmatized group would show lower bias than the general public. In addition, measures of implicit and explicit evaluation of the self as mentally ill were added to examine relationships among mental illness self-concept and mental illness biases.

METHOD

Participants

The diagnosed sample was recruited from a community mental health rehabilitation center in Virginia that provides day programs, such as daily living skills and social networking, for persons in the community who have been diagnosed with serious and persistent mental illnesses. Participants were approached at the facility and invited to participate in
the study in exchange for $7. Thirty–five clients participated in the study 2. The sample was 32% female, mean age was 43 (SD = 11.40, range = 25–76), and 77% were Caucasian (11% African American, 12% indi-

2. Approximately 25 additional individuals started participation in the study, but did not complete the study for various reasons, so their data are not reported here. For instance, they were unable to complete the IAT portion of the study and had incomplete trial data, either because of difficulties understanding the computer task, interruptions during administration at the rehabilitation center, such as an alarm sounding or a client being called, or some of the clients found it difficult to sustain attention and/or interest for the length of the IAT portion of the study, which involves repetitive classification of items. Because it is a reaction time task, it was not possible to include the data for individuals who were interrupted in some way. Thus, there may be a bias in the sample in that those clients who were less ill or less cognitively impaired were more likely to complete the procedures, and be included in the analyses.
cated “other” or did not report their ethnicity). Diagnostic information was obtained from clients’ charts with their consent (information was available for 32 out of 35 participants). Average number of diagnoses listed for clients was 1.7 (range 1 to 4) with reports indicating a lifetime history of a psychotic disorder for 59% (predominantly schizophrenia or schizoaffective disorder), mood disorder for 31% (either with or without psychotic features), personality disorder for 31% (predominantly borderline personality disorder), mental retardation or borderline intellectual functioning for 16%, substance abuse or dependence for 13%, anxiety disorders for 13%, pedophilia for 6%, and Tourette’s disorder for 3% of the clinical sample.

A control group was recruited from a local public library by approaching individuals and inviting them to participate in the study in exchange for $7. Participants who reported a history of mental illness on our screening question (included in the demographics information completed privately by participants) were excluded from analyses (N = 19), leaving a healthy control group, according to self–report, of N = 36. This sample was 39% female, mean age was 41 (SD = 14.43, range = 25 to 76), and 64% were Caucasian (19% African American, 8% Hispanic, and 8% indicated “other” or did not report their ethnicity). Chi–square tests indicated the control and diagnosed groups did not differ by gender (χ² = .33, p > .10) or race (χ² = 4.92, p > .10).

Materials

Implicit Measures. IAT tasks were identical to those used in Study 1, except that the “IAT mental illness + blameworthy” task was replaced with a new task that associated the self (using the categories “me” versus “not me”) with mental versus physical illness (referred to as “IAT mental illness + me”). See Appendix for associated stimuli.

Explicit Measures. The semantic differential items were identical to Study 1, except that the blameworthy/innocent items were replaced by a single item that asked, “I feel I am . . . ,” which required participants to rate themselves as physically versus mentally ill on a 7–point semantic differential scale. This measure was not relative, unlike the IATs and the other semantic differential scales, because we felt it would not make sense to have participants rate “others” as physically versus mentally ill.

---

3. As in Study 1, participants completed the job applicant vignette and other explicit questionnaires, but again, the vignette manipulation did not significantly affect the multivariate tests for the implicit (IAT tasks) and explicit (semantic differentials) bias measures (p > .10).
Procedure
Same as Study 1.

RESULTS

Data Reduction. IAT data were again scored according to the new scoring algorithm developed by Greenwald et al. (2003). This resulted in deletion of nine participants’ IAT data (one from the control sample and eight from the diagnosed sample) due to either high error rates and/or unusually fast or slow response times. These participants’ explicit data were retained. The extra deletion of the IAT data from the sample with mental illness likely reflects three challenges in collecting reaction time data with this sample. First, although the data collection occurred in a private room at the rehabilitation facility, the environment still had occasional interruptions. Second, it is plausible that this group had less experience working with computers, making the task more foreign, resulting in higher error rates. Third, we did not assess the sample’s current medication intake so it is not clear whether medication influenced ability to complete the task. Nonetheless, because the task relies on within-subject comparison of time to categorize stimuli, between-subjects differences in experience and the environment should not greatly affect those data that conform to the Greenwald et al. inclusion criteria.

Evidence of Implicit and Explicit Bias. Replicating the results from Study 1, there was strong evidence of an implicit bias against mental illness relative to physical illness. T-tests for the full sample (control and diagnosed groups combined) indicated that the IAT mental illness + bad ($t_{61} = 5.55, p < .0001, d = 1.42$) and the IAT mental illness + helpless ($t_{61} = 2.25, p = .03, d = .58$) were each significantly different from zero, suggesting more negative evaluations of persons with mental illness (relative to persons with physical illness). In addition, t-tests on the semantic differential scales matched the findings from Study 1, indicating relatively negative Explicit mental illness + bad ($t_{70} = 3.04, p = .003, d = .73$) and Explicit mental illness + helpless ($t_{63} = 2.35, p = .02, d = .59$) ratings. Also replicating Study 1, the explicit evaluations of the groups separately (i.e., the individual items used to calculate the difference scores) did not indicate negative evaluations of the groups, suggesting again that the bias is relative rather than absolute. Not surprisingly, the combined sample did not show stronger evaluations of themselves as mentally ill relative to physically ill on either the implicit or explicit measures (IAT mental illness + me: $t_{61} = .43, p > .10, d = .11$; Explicit mental illness + me: $t_{70} = 1.24, p > .10, d = .30$). All tests compared the $t$ value to zero with the exception of the Explicit mental illness + me, which compared the value to four
because this was the neutral point on the single item scale (since no difference score was obtained).

To compare the diagnosed and healthy control samples to one another, independent samples t-tests were conducted for each of the bias measures. Interestingly, no group differences emerged on the IAT mental illness + bad ($t_{60} = .65, p > .10, d = .17$), the IAT mental illness + helpless ($t_{60} = .84, p > .10, d = .22$), the Explicit mental illness + bad ($t_{60} = .20, p > .10, d = .05$), or the Explicit mental illness + helpless ($t_{62} = .64, p > .10, d = .16$), suggesting that persons with and without a mental illness diagnosis evaluate mentally ill persons equally negatively. More surprising, the mentally ill sample did not evaluate themselves as more mentally than physically ill compared to the control group (Explicit mental illness + me: $t_{69} = .43, p > .10, d = .10$). In fact, on the IAT mental illness + me task, the control group was implicitly higher on mental illness associations with the self ($t_{60} = 2.48, p = .02, d = .64$). This finding clearly contradicted expectations and may be due to the use of physical illness as the contrasting category to mental illness (see Discussion).

Overall, the bias measures replicated the findings from Study 1 of relative implicit and explicit biases against mental illness and in favor of physical illness, and the level of this bias did not vary as a function of one’s reported personal history of mental illness.

**Relationships Among Bias Measures and Individual Differences.** Similar to Study 1, there were few significant relationships among the bias measures. The IAT mental illness + bad and IAT mental illness + helpless tasks were positively related to one another ($r = .27, p = .03$), but neither IAT was related to IAT mental illness + me ($r = -.001$ with helpless, $r = .001$ with bad, both $p > .10$). Among the semantic differential measures, there were no significant relationships ($r$ ranged from $-.08$ to $-.19$, all $p > .10$), and only the implicit and explicit mental illness + bad measures were related ($r = .32, p = .01$). All other implicit/explicit indicators were unrelated ($r$ ranged from $-.12$ to $-.10$, all $p > .10$). Again, replicating Study 1, there were no significant gender differences on the bias measures (all $p > .10$). Finally, the Perceived Dangerousness Scale was positively related to Explicit mental illness + bad ($r = .34, p = .005$), but not to the other bias measures ($r$ ranged from $-.09$ to $-.04$, all $p > .10$), again providing partial support for the expected positive associations.

**Implicit Bias Across Samples.** Although the results across the studies cannot be compared directly, it is interesting to note what percentage of each sample demonstrated a negative versus positive implicit evaluation of persons with mental illness. For descriptive purposes, the IAT results were split according to those showing a positive IAT effect (i.e., a score above the zero-point indicating relatively negative implicit mental illness associations, which can also be interpreted as positive associa-
tions with physical illness), and those showing a negative IAT effect (below the zero-point). In Study 1, the IAT effects among the student sample were: IAT mental illness + bad (68% showed negative associations with mental illness), IAT mental illness + helpless (75% showed negative associations with mental illness), and IAT mental illness + blameworthy (58% showed negative associations with mental illness). In Study 2, among the control sample in the general public, IAT effects were: IAT mental illness + bad (78% showed negative associations with mental illness), and IAT mental illness + helpless (72% showed negative associations with mental illness). Finally, among the sample diagnosed with serious mental illness, IAT effects were: IAT mental illness + bad (80% showed negative associations with mental illness), and IAT mental illness + helpless (66% showed negative associations with mental illness). These results, while descriptive, reinforce the earlier findings of relatively similar implicit biases across the groups.

GENERAL DISCUSSION

The current study investigated implicit and explicit attitudes toward mental illness among student, general population, and diagnosed samples. Robust evidence for implicit biases that lie outside conscious control was expected given the pervasive negative messages in our culture about the incompetence (e.g., in treatment decisions; Slobogin, 1996) and culpability (Kleinke & Kane, 1997) of mentally ill persons that occur despite efforts in recent years to better inform the public and reduce mental illness stigma. These predictions were strongly supported. Implicit biases against mental illness (and in favor of physical illness) were evident both in negative attitudes and in the helplessness and blaming stereotypes. These biases were robust across samples, and both members and non-members of the stigmatized group (i.e., with or without a psychiatric diagnosis) held equally negative evaluations of persons with mental illness, supporting the pervasiveness of the biases.

The negative attitude and helplessness bias were also evident at an explicit level, suggesting that it is still relatively more socially acceptable to report negative evaluations of mentally ill persons. This finding may be partly understood in the context of the relative comparison with physical illness. The absolute evaluations for both physically and mentally ill people were generally not negative, but the findings indicate that our samples endorse more critical evaluations of mentally ill than physically ill people. With all measures of relative bias, the results can be interpreted as both anti-mental illness and/or pro-physical illness views. Interestingly, in an evaluation of the acceptability of reporting prejudice toward a broad range of social groups, Crandall, Eshleman, and O’Brien
(2002) found that college students rated the acceptability of having negative feelings about “mentally unstable people” close to the median of the 105 groups evaluated. It is noteworthy that the explicit evaluations were generally not negative toward either group, so the evidence for explicit bias can only be interpreted in comparative terms. This may suggest relatively greater tolerance today than in the past, or may be reflective of social desirability concerns to appear tolerant. In consequence, bias may be easier to observe on relative indices, such as the ones used in the current study, because evaluations of a single group may be less likely to be negative on their own, whereas comparisons across groups may reveal greater variability because of the opportunity for contrasting marginalized groups.

There was no explicit evidence of more blaming of mentally ill persons (neither relative nor absolute negativity), despite the implicit bias. One possible explanation comes from recent research suggesting that the public is now less likely to hold individuals personally responsible for their mental illness, instead citing stressful circumstances and biological causal factors more frequently (Link, Phelan, Bresnahan, Stueve, & Pescosolido, 1999). This would be an encouraging finding given evidence from attribution theory research of the negative consequences that follow from blaming an individual for a “deviant” condition (Weiner, Perry, & Magnusson, 1988), and Corrigan et al.’s (1999) finding that blaming attributions about mental illness were negatively related to anti-stigma behaviors. However, another interpretation of these findings is that given the Western emphasis on physical appearance and fitness, it is possible the public also holds people accountable for their physical health (though Corrigan et al., 2000, found that persons were held more responsible for having a mental illness relative to a physical disability). These results emphasize the need to shift public perceptions about the competence and blameworthiness of mentally ill persons (and perhaps also about physically ill persons). Further, the discrepancy between the implicit and explicit measures points to the importance of evaluating both automatic and self-reported evaluations of mental illness.

The consistency of findings across the two studies (and student, general population, and diagnosed samples) for both the implicit and ex-

---

4. The mean rating was $M = .41$ ($0 = \text{definitely not OK}$ to $2 = \text{definitely OK}$). There was no comparable general category for physically ill people, though it was considered less acceptable to report prejudice toward people with AIDS ($M = .23$; no significance test between groups was provided).
Explicit bias measures points to the robustness of the stigma, but also to the lack of in–group bias among the mentally ill participants. Contrary to expectations, there were no differences in the disparaging attitudes and beliefs about mental illness between the general public and the clinically diagnosed sample. The absence of a protective in–group bias was disappointing, because it suggests equal negativity from persons who have extensive experience with mental illness (likely in terms of both knowledge and contact, established correlates of reduced stigma; Corrigan et al., 2001) and who will be hurt most by the stigma.

There are a number of possible explanations for these findings, including the possibility that, although positive changes may be occurring explicitly (Link et al., 1999), highly blaming viewpoints remain implicitly toward persons with mental illness (matching the current study’s results). These blaming views are presumably associated with the numerous negative consequences that follow perceiving an illness as controllable (Weiner, 1993; Weiner et al., 1988), such as reduced pity or desire to help and increased anger toward the marked person. If even diagnosed persons blame themselves for becoming ill, then it is not surprising that they hold such negative views toward persons with mental illness. These findings also correspond with the strong anti–fat attitudes usually found among obese persons, which Crandall suggests may be due to the close tie between weight bias and a worldview that holds people “get what they deserve and deserve what they get” (Crandall, 1994; Crandall & Biernat, 1990). An analogous worldview may contribute to beliefs that persons with mental illness are responsible for their illness.

It is also possible that the lack of group differences follows from the absence of a mentally ill self–concept among the diagnosed sample, suggesting they may not think of themselves as part of the in–group. On the explicit measure, persons with mental illness were no more likely to associate the “self” with “mental illness” than were persons without mental illness, and most unexpectedly, on the implicit task, persons with mental illness were significantly less likely to associate the “self” with “mental illness.” One possible explanation for these findings is that persons succeeding in treatment may distance themselves from their mental illness label as a function of the rehabilitation process. It is also plausible that the mentally ill sample compared themselves to others at the rehabilitation center, and evaluated themselves as relatively less mentally ill. (This is consistent with the possibility that individuals with the least impairment were most likely to complete the procedures; see footnote 3.) Finally, the finding may be understood in the context of the relative comparison with physical illness. Perhaps the recent emphasis in the media and among various advocacy groups to describe mental illnesses as biological in origin (see Phelan, Cruz–Rojas, & Reiff, 2002, for
discussion of the impact of this emphasis) may mean the diagnosed sample thinks of themselves as physically ill as much as mentally ill. Unfortunately, the physical health of the sample was not assessed, so comorbid physical illnesses may have influenced the results. Future research that uses an alternate comparison category (besides physical illness) or non–relative measures will be helpful to make sense of these surprising findings.

It will also be important for future research to address other limitations of the current study, including using diagnosed samples that are less heterogeneous than the sample used in the current study and at different stages of illness and recovery to evaluate whether the stigma remains as negative across different disorders and levels of functional impairment. Further, it will be advantageous to use structured interviews (rather than the self–report and chart reviews used here) to screen for and diagnose mental illness in both the control and clinical samples. Finally, although the IAT is widely used in social cognition research and appears to have good psychometric properties (Greenwald & Nosek, 2001), obtaining convergent evidence from additional implicit measures of bias that are less vulnerable to the limitations of the IAT would be helpful. In particular, a non–relative task that required less dependence on speeded responses and was slightly easier to complete would increase the generalizability and interpretability of the findings because fewer participants would be excluded because of difficulties finishing the task.

CONCLUSION

Taken together, these findings present a somewhat discouraging picture. Despite recent advocacy efforts to reduce stigma, relative bias against mental illness is evident on both an implicit and explicit level, across a variety of stereotype domains, and across both healthy and diagnosed samples. Further, we found no evidence that members of the stigmatized group have more tolerant or forgiving attitudes about persons with mental illness than does the general public. Clearly, there are very positive effects that follow from mental illness labels with respect to identifying problems and guiding treatment, but society’s continued stigmatizing response to mental illness makes it one of the most marginalized conditions in modern Western societies (Link & Phelan, 1999). Further, the implicit biases make evident that even wishing to be tolerant or feeling conscious positive evaluations may not be sufficient to override the enormous number of negative social messages about mental illness encountered everyday.
REFERENCES


